UNITED STATES DISTRICT COURT SOUTHERN DISTRICT OF NEW YORK

IN RE:

OPENAI, INC.,

COPYRIGHT INFRINGEMENT LITIGATION

25-md-3143 (SHS) (OTW)

This Document Relates To:

No. 1:23-cv-11195 (SHS) (OTW)

NEWS PLAINTIFFS' SUPPLEMENTAL MEMORANDUM OF LAW REGARDING OPENAI'S DELETION OF OUTPUT LOG DATA

TABLE OF CONTENTS

I.	INTRODUCTION			
II.	FACTUAL BACKGROUND			
	A.	The Importance of OpenAI's Output Log Data	3	
	B.	The Parties' Meet and Confer Efforts	4	
III.	LEGAL STANDARDS			
IV.	ARGUMENT			
	A.	OpenAI Had a Duty to Preserve Relevant Output Log Data No Later than the Filing of The Times's Original Complaint on December 27, 2023	6	
	B.	News Plaintiffs are Prejudiced by OpenAI's Significant, Ongoing Destruction of Output Log Data	8	
V.	CONCLUSION		10	

TABLE OF AUTHORITIES

•	a	C	Δ
•	а		L.

Arista Records, Inc. v. Mp3Board, Inc., No. 00-cv-4660, 2002 WL 1997918 (S.D.N.Y. Aug. 29, 2002)	9
Columbia Pictures Indus. v. Bunnell, No. 06-cv-1093, 2007 WL 2080419, (C.D. Cal. May 29, 2007)	8
Concord Music Grp., Inc. v. Anthropic PBC, No. 24-cv-03811-EKL (N.D. Cal. Mar. 25, 2025)	2
Fujitsu Ltd. v. Federal Exp. Corp., 247 F.3d 423 (Fed. Cir. 2001)	6
Int'l Swaps & Derivatives Ass'n, Inc. v. Socratek, L.L.C., 712 F. Supp. 2d 96 (S.D.N.Y. 2010)	9
Ransom v. Andrews, No. 21-cv-6343, 2022 WL 16555362 (S.D.N.Y. Oct. 31, 2022)	6
Simmons v. Danhauer & Assocs., LLC, No. 08-cv-3819, 2009 WL 10677391 (D.S.C. Sept. 15, 2009)	8
Viacom Int'l Inc. v. Youtube Inc., 253 F.R.D. 256 (S.D.N.Y. 2008)	8
Zubulake v. UBS Warburg LLC, 220 F.R.D. 212 (S.D.N.Y. 2003)	6
Rules	
Fed. R. Civ. P. 37(e)	6

The New York Times Company ("The Times"), Daily News, LP, et al. ("Daily News Plaintiffs"), and The Center for Investigative Reporting ("CIR") (collectively, "News Plaintiffs") submit this supplemental memorandum pursuant to this Court's May 13 Order. <u>Dkt. 551</u>.¹

I. INTRODUCTION

As OpenAI recognized at the very inception of The Times's lawsuit: "Plaintiff's complaint focuses heavily on the 'outputs of Defendants' GenAI models." Dkt. 72 at ¶ 41. But after significant discovery efforts into OpenAI's output log data, and nearly eleven months after The Times filed its Complaint, News Plaintiffs learned that OpenAI was destroying relevant output log data. News Plaintiffs also learned only very recently that the fraction of ChatGPT Free, Pro, and Plus output log data that OpenAI has destroyed is quite substantial. Based on this information, News Plaintiffs submit that spoliation sanctions may be warranted.

First, despite OpenAI's duty to preserve all relevant output log data due to its centrality to this case, it failed to take any steps to suspend its routine destruction practices—except with respect to a handful of user accounts that OpenAI believes were involved in generating the outputs cited in the complaints (that is, output log data that OpenAI thought would be helpful to its case). Although News Plaintiffs have made repeated attempts to work with OpenAI to address its concerns about user privacy—notwithstanding that OpenAI's own privacy policy specifically advises customers that its retention of output log data is "subject to" any legal requirements to

¹ Unless otherwise indicated, all docket reference are to the docket in *The New York Times*, Case No. 1:23-cv-11195.

² With respect to OpenAI's argument that it had first informed The Times that it was not retaining certain output data on February 29, 2024 rather than on November 15, 2024, that issue was discussed in detail at the January 22, 2025 hearing. Ex. A at 46:19-49:10. Once Mr. Crosby emailed OpenAI on March 5, 2024 (<u>Dkt. 379-4</u>) advising OpenAI that The Times expected OpenAI to comply with its preservation obligations, The Times assumed that OpenAI would do so. It was not until OpenAI's November 15, 2024 letter that News Plaintiffs learned that OpenAI had chosen instead to destroy output log data.

³ Although OpenAI has not identified the fraction of ChatGPT Enterprise user conversations and API Platform completions that OpenAI has destroyed, News Plaintiffs believe the fraction of destroyed data for these products is even greater in view of **OpenAI's default retention policy to discard such data after 30 days**. *Enterprise privacy at OpenAI*, OPENAI (Oct. 31, 2024), https://openai.com/enterprise-privacy ("[a]fter 30 days, API inputs and outputs are removed from our systems, unless we are legally required to retain them").

retain such data (*see* note 10, *infra*), OpenAI has refused to suspend its destruction practices. In contrast, Microsoft appears to have complied with its discovery obligations.⁴

Second, News Plaintiffs are prejudiced by OpenAI's ongoing destruction of output log data. It diminishes News Plaintiffs' ability to prove their output-based claims by destroying evidence of the use of News Plaintiffs' content in OpenAI's products and how users employ those products in a substitutive way. It also impedes News Plaintiffs' ability to rebut OpenAI's assertions that "regurgitations" and "hallucinations" are "uncommon and unintended behaviors," Daily News Dkt. 82 at 6, or prove that OpenAI knew about them, which is relevant to scienter under the DMCA. This prejudice is exacerbated because OpenAI

It is also reasonable to assume (as this Court hypothesized on January 22, 2025)⁵ that users who were using ChatGPT to evade publishers' paywalls were also more likely to ask OpenAI to delete their output data, and

therefore the prevalence of harmful outputs is higher in the destroyed data.

Third, OpenAI's conduct suggests an intent to deprive News Plaintiffs of relevant output data for use in the litigation. OpenAI has delayed and obstructed News Plaintiffs' efforts during the past six months to obtain the most basic information about OpenAI's destruction practices. In order to assess the full degree of prejudice that News Plaintiffs have suffered, which is relevant to potential spoliation sanctions, News Plaintiffs ask this Court to: (a) keep in place its interim

⁴ Another GenAI company, Anthropic, also appears to have complied with its discovery obligations. *See Concord Music Grp., Inc. v. Anthropic PBC*, No. 24-cv-03811-EKL, Dkt. No. 318, slip op. at 2 (N.D. Cal. Mar. 25, 2025). ⁵ "The Court: [H]ypothetically, say, a ChatGPT user who had been using, found some way to get around the pay wall, right, and was getting The New York Times content somehow as the output, found a way to do it. And then hears about this case and says, Oh, woah, you know, I'm going to ask them to delete all of my searches and not retain any of my searches going forward." Ex. A at 38:14-20.

preservation order (<u>Dkt. 551</u>); (b) order OpenAI to immediately provide targeted discovery to determine the extent of prejudice caused by OpenAI's destruction of output data; and (c) order OpenAI to perform a search over the full extant output data for News Plaintiffs' content.

II. FACTUAL BACKGROUND

For nearly a year, News Plaintiffs have been trying, without success, to obtain full production of "output log data" for Defendants' products at issue in the litigation. Although OpenAI provides some details about the volume of its destruction of output log data for ChatGPT Free, Pro, and Plus in its March 31, 2025 letter (<u>Dkt. 546-2</u>), OpenAI has provided no information about the volume of destroyed data for the API Platform and ChatGPT Enterprise.

A. The Importance of OpenAI's Output Log Data

OpenAI's output log data for third-party users is highly relevant to the issues in the News cases. As Judge Stein correctly recognized, News Plaintiffs allege several theories of liability pertaining to the outputs of Defendants' products: (1) regurgitations of News Plaintiffs' works, (2) "hallucinations . . . that misattribute content to plaintiffs that they did not in fact publish," and (3) retrieval augmented generation ("RAG") responses that include "paraphrases and direct quotes of plaintiffs' works, without referring users to plaintiffs' websites in the same manner as do regular internet search engines, thereby obviating the need for users to visit plaintiffs' websites." Dkt. 514 at 8-9. For the DMCA claim, Judge Stein held that News Plaintiffs also plausibly alleged OpenAI's knowledge, or reason to know, that removing CMI would induce or conceal infringement because the models are "capable of distributing unlicensed copies of copyrighted works." Dkt. 514 at 27.

_

⁶ Output log data includes the prompts and outputs of OpenAI's products, and OpenAI's products include: (1) ChatGPT, which includes Browse, SearchGPT, and CustomGPTs such as "Remove Paywall" (collectively, "ChatGPT"); and (2) the application programming interface platform ("API Platform"). <u>Dkt. 301</u> at 1.

To prevail on these claims, News Plaintiffs seek access to OpenAI's output log data to show, inter alia, the output logs contain infringing copies of News Plaintiffs' copyrighted works, responses that are based on and are substitutive of such copyrighted works, hallucinated content that dilutes the News Plaintiffs' famous trademarks, and that OpenAI knew, or had reason to know, that removing CMI would facilitate unlicensed outputs of News Plaintiffs' content. Even OpenAI has admitted that ChatGPT outputs are "potentially infringing." See Ex. B at 10.

Defendants' output log data is also particularly germane to the first factor (transformative use) and fourth factor (effect on the potential market) of the fair use analysis. Ex. C at 47 ("The use of RAG is less likely to be transformative where the purpose is to generate outputs that summarize or provide abridged versions of retrieved copyrighted works, such as news articles, as opposed to hyperlinks."), 65, 73; Ex. D at 69:11-15 ("[I]f there's a regurgitation of 25 percent of a news article, the heart of the news article or whatever, then the fact that the other 75 percent is not regurgitated . . . does not matter and does not help the defendant, right?"), 9-10, 37-38.

B. The Parties' Meet and Confer Efforts

The parties' correspondence up through January 13, 2025 is summarized in the parties' earlier briefing and was previously addressed at the January 22, 2025 hearing. Dkts. 379, 420; Ex. A at 46:19-49:10. Since then, OpenAI has maintained its position that the size of the data and privacy and contractual obligations with its end users prohibit it from adhering to its preservation obligations in these lawsuits. Despite News Plaintiffs' efforts to work with OpenAI to come up with a technical solution for identifying and retaining a smaller set of relevant outputs to address these concerns, OpenAI has represented to News Plaintiffs, and to this Court, that it has no way of identifying relevant conversations for preservation. See Dkt. 423 at 2. But as noted in News Plaintiffs' February 5, 2025 letter to OpenAI, this representation is belied by deposition testimony from Mike Trinh, OpenAI's Associate General Counsel, who testified that OpenAI

2; see also Ex. F (Trinh Dep. Tr.) at 96:6-98:6; 103:20-104:11. In other words, OpenAI

Despite News Plaintiffs' numerous requests dating back to December 10, 2024 that OpenAI identify the volume of output log data it destroyed—which culminated in several rounds of letter correspondence, two motions to compel, and a 30(b)(6) custodial deposition during which OpenAI's witness could not even estimate the approximate volume of destroyed output log data (see Ex. F at 106:8-14)—OpenAI delayed doing so until it was ordered by the Court to respond by March 31, 2025. Dkt. 484. And OpenAI still has not provided a complete accounting.

In parallel, and notwithstanding the substantial destruction of output log data that has occurred, News Plaintiffs have undertaken extensive efforts to work with OpenAI to learn more about the preserved output log data and develop a strategy for searching and sampling that data. After months of delays, OpenAI finally participated in a technical discussion with News Plaintiffs' experts on April 9, 2025, but presented only a single third-party consultant who did not have the necessary knowledge of OpenAI's output log data. Ex. G.

On April 11, 2025, OpenAI produced 3 terabytes of information about the output log data it has preserved for ChatGPT Free, Pro, and Plus. After comparing OpenAI's retained ChatGPT output log data to the volume of deleted data, News Plaintiffs filed a supplemental letter regarding OpenAI's destruction of output log data on May 7, 2025. Dkt. 545.

III. LEGAL STANDARDS

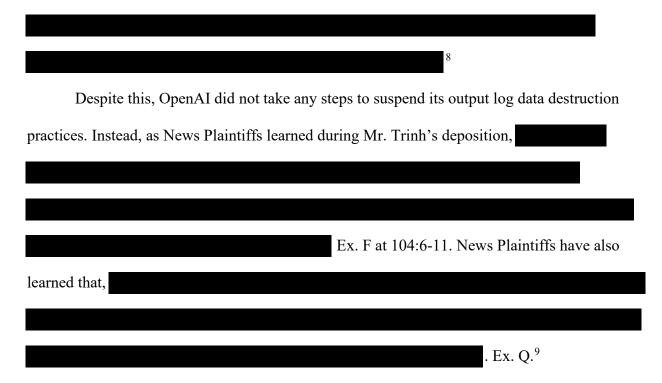
Rule 37(e) provides for sanctions where a party "failed to take reasonable steps to preserve" electronically stored information (ESI) "that should have been preserved in the anticipation or conduct of litigation," and where "it cannot be restored or replaced through additional discovery." Fed. R. Civ. P. 37(e). The Court must determine the following to award sanctions under Rule 37(e): "(1) whether a party failed to take reasonable steps to preserve ESI that should have been preserved in the anticipation of litigation; (2) whether there has been prejudice to another party from the loss of the ESI ... and (3) whether the party responsible for the spoliation acted with the intent to deprive another party of the information's use in the litigation." Ransom v. Andrews, No. 21-cv-6343, 2022 WL 16555362, at *3 (S.D.N.Y. Oct. 31, 2022) (citations and internal quotation marks omitted); Zubulake v. UBS Warburg LLC, 220 F.R.D. 212, 218 (S.D.N.Y. 2003).

IV. **ARGUMENT**

A. OpenAI Had a Duty to Preserve Relevant Output Log Data No Later than the Filing of The Times's Original Complaint on December 27, 2023

OpenAI should have taken reasonable steps to preserve relevant output log data no later than December 27, 2023, the filing and service date of The Times's Original Complaint. In fact, OpenAI's obligation to preserve relevant output log data arose as early as April 2023 when The Times first reached out to OpenAI "to raise intellectual property concerns" about OpenAI's use of Times content (Times Original Compl. at ¶ 54 (Dkt. 1)), because at that point OpenAI "should have known that the evidence may be relevant to future litigation." Fujitsu Ltd. v. Federal Exp. Corp., 247 F.3d 423, 436 (Fed. Cir. 2001). OpenAI's own internal documents dated around that time reflect that OpenAI not only knew that ChatGPT could ⁷ but that

⁷ See Ex. H (OPCO NEWS 0841405) at -410.



OpenAI contends that it was not required to preserve output logs because the data is voluminous and there are privacy considerations for its end-users. *See* Dkt. 558. They are wrong for at least three reasons. *First*, OpenAI has offered no explanation as to why it would be a significant burden to segregate and store the additional output log data that it otherwise would have destroyed, particularly given the vast computing/data storage resources at their disposal. *Second*, notwithstanding News Plaintiffs' (and the Court's) efforts to address OpenAI's privacy concerns, including News Plaintiffs' representation that they are not seeking personally

⁹ OpenAI has since clawed back all but one of the shield its cover up efforts on the basis of privilege. News Plaintiffs have asked for a meet and confer.

retention of output log data is "subject to" any legal requirements to retain such data. 10

Courts routinely reject arguments that a privacy policy supersedes a party's obligation to preserve and produce information relevant to a litigation. See Viacom Int'l Inc. v. Youtube Inc., 253 F.R.D. 256, 262 (S.D.N.Y. 2008) (rejecting arguments over "users' privacy concerns" and ordering defendant to produce user log data); Columbia Pictures Indus. v. Bunnell, No. 06-cv-1093, 2007 WL 2080419, at *8 (C.D. Cal. May 29, 2007) (rejecting privacy policy arguments and ordering party to retain and produce server logs); Simmons v. Danhauer & Assocs., LLC, No. 08-cv-3819, 2009 WL 10677391, at *2 (D.S.C. Sept. 15, 2009) (finding "no violation of the privacy policy could reasonably ensue" where privacy policy permits disclosure "when required by law"). Third, similarly situated defendants have figured out how to preserve output log data for use in litigation—including Microsoft—and there is no reason why OpenAI, as one of the most sophisticated technology companies in the world, could not also solve this problem.

B. News Plaintiffs are Prejudiced by OpenAI's Significant, Ongoing Destruction of Output Log Data

OpenAI's output log data for all products at issue is indisputably relevant to News Plaintiffs' claims and Defendants' defenses, and News Plaintiffs are prejudiced by being prevented from examining the significant volume of output log data that OpenAI destroyed.

In order to support their output-based claims, News Plaintiffs have propounded discovery on OpenAI seeking output data from its products that contain, *inter alia*, infringing reproductions

_

¹⁰ See, OpenAI, Privacy Policy, OPENAI, https://openai.com/policies/row-privacy-policy/ ("How long we retain personal Data will depend on a number of factors, such as: . . . Any legal requirements that we are subject to.") (emphasis added); OpenAI, Enterprise privacy at OpenAI, OPENAI, https://openai.com/enterprise-privacy/ ("Any deleted or unsaved conversations are removed from our systems within 30 days, unless we are legally required to retain them.") (emphasis added); OpenAI, Data Controls in the OpenAI platform, OPENAI, https://platform.openai.com/docs/models/how-we-use-your-data ("API data may be retained for up to 30 days, after which it will be deleted (unless otherwise required by law)") (emphasis added).

of News Plaintiffs' copyrighted works or hallucinations that dilute News Plaintiffs' famous marks. *See, e.g., Arista Records, Inc. v. Mp3Board, Inc.,* No. 00-cv-4660, 2002 WL 1997918, at *3 (S.D.N.Y. Aug. 29, 2002) (Stein, J.); *Int'l Swaps & Derivatives Ass'n, Inc. v. Socratek, L.L.C.*, 712 F. Supp. 2d 96, 102 (S.D.N.Y. 2010) (Stein, J.). OpenAI's destruction of a significant portion of its output data hamstrings News Plaintiffs from showing that those outputs infringe News Plaintiffs' copyrighted works, dilute News Plaintiffs' trademarks, or evidence OpenAI's scienter in removing News Plaintiffs' CMI. Moreover, it is entirely reasonable to infer that the prevalence of infringement may be higher in the destroyed data and that sampling the extant logs will not cure the prejudice caused by OpenAI's destruction. Users engaged in infringing behavior, including circumventing paywalls or generating "pink-slime" journalism by copying the News Plaintiffs' content, are far more likely to instruct OpenAI to destroy their conversation data to cover their tracks.

Finally, as part of its defensive themes, OpenAI has repeatedly represented that people do not interact with its products in the "real world" in a manner that infringes copyright, and that regurgitations and output based intellectual property violations are exceedingly rare. ¹¹ OpenAI's destruction of relevant output log data impedes News Plaintiffs' ability to refute OpenAI's representations that its models rarely memorize and regurgitate training data, including News Plaintiffs' copyrighted content. Moreover, OpenAI's representations are belied both by the documents that Defendants' have produced, *see* note 8, *supra*, and the affirmative steps OpenAI has taken

_

¹¹ See, e.g., <u>Dkt. 52</u> at 1 ("...ChatGPT is not in any way a substitute for a subscription to The New York Times. In the real world, people do not use ChatGPT or any other OpenAI product for that purpose."); *Daily News* <u>Dkt. 82</u> at 2 ("...Plaintiffs have flagrantly mischaracterized what OpenAI's products are and how they work."); *id.* at 6 ("As to the outputs, Plaintiffs focus on two uncommon and unintended behaviors that they say harm them in distinct ways: regurgitation and hallucination."); Ex. B at 10 (OpenAI's October 23, 2023 Initial Comments to USCO Notice of Inquiry) ("[D]ue to steps taken during the pre-training process (including in collaboration with rightsholders) and the design of our models, memorization of training data is exceptionally rare.").

. OpenAI's

are also relevant to News Plaintiffs' copyright,

DMCA and willful infringement claims. OpenAI's output logs from before and immediately after The Times filed suit—

12—contain evidence essential to investigate OpenAI's

V. CONCLUSION

To assess the full degree of prejudice that News Plaintiffs have suffered, News Plaintiffs ask this Court to: (a) keep in place its interim preservation order (Dkt. 551); (b) order OpenAI to promptly provide a declaration and relevant documents, and make available for deposition one or more knowledgeable 30(b)(6) witnesses, on the following topics: (1) the full scope of destroyed output log data for all of its products at issue; (2) any efforts by OpenAI to ascertain whether News Plaintiffs' intellectual property was present in, or was the subject of a request by a user, in the destroyed output data; (3) any alternative sources of relevant information OpenAI has regarding the destroyed output log data; (4) whether the destroyed output log data containing News Plaintiffs' content can be retrieved; and (5) the and (c) order OpenAI to perform a search over the full extant output log data for News Plaintiffs' content.

¹² OpenAI's Letter Motion Seeking Reconsideration of this Court's Order (the "Reconsideration Motion") provides no explanation for why

Further, though OpenAI's Reconsideration Motion claims that

Dated: May 16, 2025 /s/Ian Crosby

Ian Crosby (pro hac vice)
Genevieve Vose Wallace (pro hac vice)
Katherine M. Peaslee (pro hac vice)
SUSMAN GODFREY L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 516-3880
Facsimile: (206) 516-3883
icrosby@susmangodfrey.com
gwallace@susmangodfrey.com
kpeaslee@susmangodfrey.com

Davida Brook (pro hac vice)
Emily K. Cronin (pro hac vice)
Ellie Dupler (pro hac vice)
SUSMAN GODFREY L.L.P.
1900 Ave of the Stars, Suite 1400
Los Angeles, CA 90067
Telephone: (310) 789-3100
Facsimile: (310) 789-3150
dbrook@susmangodfrey.com
ecronin@susmangodfrey.com
edupler@susmangodfrey.com

Elisha Barron (5036850)
Zachary B. Savage (ZS2668)
Tamar Lusztig (5125174)
Alexander Frawley (5564539)
Eudokia Spanos (5021381)
SUSMAN GODFREY L.L.P.
One Manhattan West
New York, NY 10001
Telephone: (212) 336-8330
Facsimile: (212) 336-8340
ebarron@susmangodfrey.com
zsavage@susmangodfrey.com
tlusztig@susmangodfrey.com
afrawley@susmangodfrey.com
espanos@susmangodfrey.com

Scarlett Collings (4985602) SUSMAN GODFREY L.L.P. 1000 Louisiana, Suite 5100 Houston, TX 77002 Telephone: (713) 651-9366 Facsimile (713) 654-6666 scollings@susmangodfrey.com

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Kristen J. Logan (pro hac vice)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
901 New York Avenue, N.W., Suite 900 East Washington, DC 20001
Telephone: (202 783-6040
Facsimile: (202) 783 6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com
klogan@rothwellfigg.com

Attorneys for Plaintiff The New York Times Company

/s/Steven Lieberman Steven Lieberman (SL8687) Jennifer B. Maisel (5096995) Robert Parker (pro hac vice) Jenny L. Colgate (pro hac vice) Mark Rawls (pro hac vice) Kristen J. Logan (pro hac vice) Bryan B. Thompson (6004147) Mary L. Mullin (pro hac vice) ROTHWELL, FIGG, ERNST & MANBECK, P.C. 901 New York Avenue, N.W., Suite 900 East Washington, DC 20001 Telephone: (202) 783-6040 Facsimile: (202) 783-6031 slieberman@rothwellfigg.com jmaisel@rothwellfigg.com rparker@rothwellfigg.com icolgate@rothwellfigg.com mrawls@rothwellfigg.com klogan@rothwellfigg.com

Jeffrey A. Lindenbaum (JL1971) ROTHWELL, FIGG, ERNST &

bthompson@rothwellfigg.com mmullins@rothwellfigg.com

MANBECK, P.C.

3 Manhattanville Road, Suite 105 Purchase, New York 10577 Telephone: (202) 783-6040 Facsimile: (202) 783-6031 jlindenbaum@rothwellfigg.com

Attorneys for Plaintiffs Daily News, LP; The Chicago Tribune Company, LLC; Orlando Sentinel Communications Company, LLC; Sun-Sentinel Company, LLC;

San Jose Mercury-News, LLC; DP Media Network, LLC; ORB Publishing, LLC; and

Northwest Publications, LLC

/s/Matthew Topic

Jonathan Loevy (pro hac vice) Michael Kanovitz (pro hac vice) Lauren Carbajal (pro hac vice) Stephen Stich Match (No. 5567854) Matthew Topic (pro hac vice) Thomas Kayes (pro hac vice) Steven Art (pro hac vice) Kyle Wallenberg (pro hac vice)

LOEVY & LOEVY

311 North Aberdeen, 3rd Floor Chicago, IL 60607 312-243-5900 (p) 312-243-5902 (f) jon@loevy.com mike@loevy.com carbajal@loevy.com match@loevy.com matt@loevy.com steve@loevy.com kayes@loevy.com wallenberg@loevy.com

Attorneys for Plaintiff The Center for Investigative Reporting, Inc.